The Dynamics of Plant-level Productivity in U.S. Manufacturing

Árpád Ábrahám*

Kirk White†

April 4, 2005

Abstract

Using a unique database that covers the entire U.S. manufacturing sector from 1976 until 1999, we estimate plant-level total factor productivity for a large number of plants. We characterize time series properties of plant-level idiosyncratic shocks to productivity, taking into account aggregate manufacturing-sector shocks and industry-level shocks. Plant-level heterogeneity and shocks are a key determinant of the cross-sectional variations in output. We compare the persistence and volatility of the idiosyncratic plant-level shocks to those of aggregate productivity shocks estimated from aggregate data. We find that the persistence of plant level shocks is surprisingly low-we estimate an average autocorrelation of the plant-specific productivity shock of only 0.37 to 0.41 on an annual basis. Finally, we find that estimates of the persistence of productivity shocks from aggregate data have a large upward bias. Estimates of the persistence of productivity shocks in the same data aggregated to the industry level produce autocorrelation estimates ranging from 0.80 to 0.91 on an annual basis. The results are robust to the inclusion

*University of Rochester, Department of Economics. Contact: aabraha2@mail.rochester.edu. Phone: 585-275-5273 Fax: 585-256-2309

†U.S. Bureau of the Census. Contact: Kirk.White@duke.edu. Phone: 919-660-6893. Fax: 919-684-8974. The research in this paper was conducted while the second author was an employee of the U.S. Census Bureau at the Triangle Census Research Data Center. Research results and conclusions expressed are those of the authors and do not necessarily reflect the views of the Census Bureau. This paper has been screened to insure that no confidential data are revealed. We thank participants at the Society for Economic Dynamics Meetings, and Society for Computational Economics conference, the Census Bureau Center for Economic Studies' Seminar, the Annual Census CES-RDC Meetings in Ann Arbor, Michigan, and the Duke Applied Microeconomics Lunch Group for helpful comments.

of various measures of lumpiness in investment and job flows, different weighting methods, and different measures of the plants' capital stocks.

JEL Classification codes: D24, L6, O47

Keywords: productivity, manufacturing, microdata

1 Introduction

Recent work in I.O. has emphasized the importance of firm- and plant-level heterogeneity in total factor productivity. Jensen and McGuckin (1996 CES working paper) argue that the major empirical regularity in studies of firm or establishment-level productivity is heterogeneity within sectors and across plant characteristics. They argue that economists must move beyond the representative firm model of industry-level productivity (not to mention economy-level productivity).

This paper seeks to apply the insights of the literature on idiosyncratic shocks to inividual labor productivity[1] to the dynamics of plant-level total factor productivity. We estimate establishment-level productivity using a unique longitudinal data set that covers covers the entire U.S. manufacturing sector from 1976 until 1999. Then we characterize the time series properties of establishment-level idiosyncratic shocks to productivity, taking into account aggregate economy-wide and industry-level shocks. We compare the persistence and volatility of the idiosyncratic shocks to the persistence and volatility of aggregate productivity shocks estimated from aggregate data.

This research contributes to the literature in two ways. First, much of the empirical literature on productivity focuses on a single industry or a small number of industries. While careful single-industry studies are useful, both microeconomists and macroeconomists would be interested in consistent measures of plant-level productivity across

[1] See, for example, Storesletten, Telmer and Yaron (1999).

industries. This study provides these measures.

Second, many studies of the dynamics of productivity focus on measuring year-by-year or multi-year average productivity within an industry. This research augments that approach by estimating an establishment-level idiosyncratic productivity process. There are at least two reasons for macroeconomists to be interested in this. First, it is interesting to see, aggregating from our establishment level productivity estimates, if the dynamics of aggregate productivity are the same as the dynamics of aggregate productivity estimated from aggregate data. Second, it is informative to compare estimates of the persistence and volatility of idiosyncratic productivity shocks to the estimated persistence and volatility and of aggregate productivity shocks. A priori, one would expect the idiosyncratic process to be different from the aggregate process for three reasons: (i) the aggregate process should be less volatile than the idiosyncratic process (high and low shocks in the idiosyncratic process cancel out); (ii) because of the exit of less productive firms; and (iii) the aggregate measure of productivity will put more weight on firms with larger stocks of capital.

We estimate plant-level AR(1) processes with plant-level fixed effects. Consistent with the existing literature, we find that heterogeneity across plants is enormous in terms of the underlying process for their shocks. We also find that the average persistence of plant level shocks is surprisingly low-we estimate an average autocorrelation of the plant-level productivity shock of only 0.37 to 0.41 on an annual basis. We also find that estimates of the persistence of productivity shocks from aggregate data have a large upward bias. Estimates of the persistence of productivity shocks in the same data aggregated to the industry level produce autocorrelation estimates ranging from 0.80 to 0.91 on an annual basis. The results are robust to the inclusion of various measures of lumpiness in investment and job flows, different weighting methods, different sample selection methods, and different measures of the plants' capital stocks.

2 The Data

We use the U.S. Census Bureau's Longitudinal Research Database (LRD) and the Bureau's Longitudinal Business Database (LBD).[2] The LBD currently runs from 1976 to 1999 and is described in Jarmin and Miranda (2002). The LBD provides the best available longitudinal links for plants across a broad range of industries. The LRD consists of the Census of Manufactures (CM), which is conducted in years ending in 2 and 5, and the Annual Survey of Manufactures (ASM) files, which are available in non-census years. The LRD contains plant-year observations of production workers hours, production and non-production workers wages, the book value of the plant's equipment and structures capital stock, capital expenditures, the cost of materials inputs, and the value of shipments. For our purposes, the main virtues of the LRD are that (i) it is a large, statistically representative, longitudinal sample of U.S. manufacturing plants; and (ii) large plants are included in practically every year. The main disadvantage of the LRD for us is that because of the sampling strategy of the ASM, smaller plants often appear for only 5 years. Our sample and the measures we use are described in more detail in the Appendix.

3 Related Literature

The LRD has been used increasingly in recent years to study plant level productivity. This literature has been reviewed by Bartelsman and Doms (2000). Two of the main findings are that there is a great deal of heterogeneity of productivity at the plant level, and that differences in productivity across plants are very persistent. Dwyer (1996) also uses the LRD to study the persistence of plant-level productivity in a 16-year panel of textile plants. He finds that statistical tests using a standard parametric approach

[2]The datasets are available to qualified researchers via the the Census Research Data Center network. See the Census' Center for Economic Studies website (http:/www.ces.census.gov) for more information.

fail to reject an hypothesis of plant-level fixed effects. However, using a nonparametric approach, he finds that differences in plant productivity have a half-life of approximately 10 to 20 years.

Dynamic and (ex post) heterogeneous firm literature. (Jovanovic (1982), Ericson and Pakes (1995), Cooley and Quadrini (2001).)

Two empirical questions that come out of this theoretical literature:

- How should we calibrate the technological process?
- Does the aggregate Solow residual give a reasonable estimate for plant-level technological progress?

4 Methodology

We use a reduced form estimation approach. Other estimation strategies (e.g., Olly and Pakes (1996); Levinsohn and Petrin (1999)) do a better job of taking account of the endogeneity of capital or of selection bias because of exit. However, the reduced form approach has several advantages. The first is comparability: we want to compare our estimates to the macro literature, and much of macroeconomic literature on productivity is reduced form. We can compare different sectors and we are able to recover industry-specific demand shocks and a clean measure of the aggregate "technology shock." We are estimating production functions for over 400 industries and productivity processes for thousands of plants. So a second advantage of a reduced form approach is that it is less computationally costly than Olley and Pakes' method.

Because we do not explicitly take account of demand shocks, the residuals we estimate are perhaps more properly called "measured TFP" rather than "productivity" shocks. However, we believe that we have uncovered some interesting and novel facts. Moreover, these are facts that any model of plant-level dynamics needs to explain.

5 The model

We estimate the model in equation (1) separately for each of 453 4-digit SIC industries. We assume that plants use a Cobb-Douglas technology; productivity is Hicks-neutral and multiplicative; and production function parameters are constant across plants and over time for a four-digit industry, but possibly different across industries.

$$\begin{aligned}\log(y_{ijt}) &= \log(A_{ijt}) + \alpha_i \log(k_{ijt}) + \beta_i \log(l_{ijt}) + \sum_{k=0}^{2} \gamma_{ik} I_{ijt}^{age=k} J^{k=t-1976} \\ &\quad + \zeta_i age_{ijt} \qquad (1)\end{aligned}$$

The subscript i is the 4 digit SIC industry; j is the plant; and t is the year. y is real value added; A is productivity; k is the plant's capital stock, which we construct from the book value and capital expenditure using the perpetual inventory method. l is a measure of production-worker-equivalent hours. AGE is a measure of the plant's age. There is a more detailed explanation of our measures in the data appendix. $I^{age=k}$ is an indicator variable that is 1 when the plant is age k and zero otherwise. This is meant to capture the idea that a plant may have different productivity in its early years, possibly because of setup costs. The J indicator variable has the effect of dropping these "young plant" dummies in the first three years of the sample, since by construction all plants that appear in the first year of the sample are age 0 in 1976, age 1 in 1977 and age 2 in 1978.

Productivity A is our main focus, and we break it into several components:

$$\log(A_{ijt}) = \log(\theta_t) + \log(\eta_{it}) + \log(z_{ijt}) \qquad (2)$$

In equation (2), θ_t measures aggregate shocks (to technology and/or policy). The parameter η_{it} captures 4-digit industry-specific shocks, which again could be shocks to technology or policy. Plant-specific shocks to productivity are measure by z_{ijt}.

We decompose the plant-specific shocks into three parts:

$$\log(z_{ijt}) = (1-\rho_j)\widetilde{\log(z_j)} + \rho_j \log(z_{ijt-1}) + \sigma_j \varepsilon_{jt} \tag{3}$$

where ε_{jt} is i.i.d. $N(0,1)$.

A plant "fixed" effect is captured by $\tilde{z}_j$. (We estimate both $\tilde{z}_j$ and $\widetilde{\log(z_j)}$.) The persistence of the plant-specific productivity shock is given by ρ_j. The parameter σ_j gives the standard deviation of the plant-specific shock.

6 Estimation

The estimation proceeds in two stages. In the first stage we estimate the industry production function separately for each industry i:

$$\begin{aligned}\log(y_{ijt}) \quad &= \quad \sum_{k=1977}^{1999} \lambda_{it} I_{it}^{YEAR=k} + \alpha_i \log(k_{ijt}) + \beta_i \log(l_{ijt}) \\ &\quad + \sum_{k=1}^{3} \gamma_{ik} I_{ijt}^{age=k} + \zeta iage_{ijt} + u_{ijt} + constant \end{aligned} \tag{4}$$

The residual u_{ijt} is an estimate of the log of plant-specific productivity shocks, $\log(z_{ijt})$. The coefficients λ_{it} on the year dummy variables are (non-parametric) estimates of $\log(\theta_t) + \log(\eta_{it})$. 1976 is the ommitted dummy.

In the second stage of the estimation, we take the residuals u_{ijt}, and estimate ρ_j and $\widetilde{\log(z_j)}$ in equation (3) plant by plant using OLS. Finally, we estimate σ_j from the residual of the above regression. Note that one can estimate $\log(\theta_t)$ as the weighted mean of λ_{it}. Then $\log(\eta_{it}) = \lambda_{it} - \log(\theta_t)$.

7 Results

7.1 First stage

Our industry coding scheme gives us 460 industries. We have enough observations to estimate equation (4) for 453 industries. The data appendix describes our industry coding scheme in detail.

Figure 1 shows a plot of the production function parameters from the first stage of the estimation. Typically the production function parameter estimates are highly statistically significant. The key points to notice from the graph are that there is huge variation in the parameter estimates across industries, and most industries seem to be constant returns to scale. When we looked at individual industries, between-industry variations typically could be rationalized by differences in capital- or labor-intensity of different industries. The shipments-share-weighted means of α and β are 0.29 and 0.69, respectively.[3]

7.2 Second stage results: plant level

In the first stage of estimation we were able to use all plants within an industry to estimate equation (1). In the second stage of estimation, we encounter a data problem. In order to estimate the three parameters of the $AR(1)$ with a plant fixed effect, we need to have a long enough time series for each plant. So we estimated equation (3) only for plants with > 15 observations. This reduces the sample from 191,200 to 19,552 plants. The good news is that on average these plants produce 61% of manufacturing value of

[3] In an attempt to account for the endogeneity of capital, we also estimated equation (4) using the lagged value of the log of capital. Theory predicts that the coefficient on current capital is downward biased because of the endogeneity of capital. However, when we use lagged capital we find that the shipments-share-weighted average estimate of α is 0.20, somewhat lower than the average when we used current capital. The average estimate of β, 0.78, is higher than the average coefficient on labor when we used current capital. Using lagged capital decreases our sample size somewhat since we have to throw out the first observation for each plant.

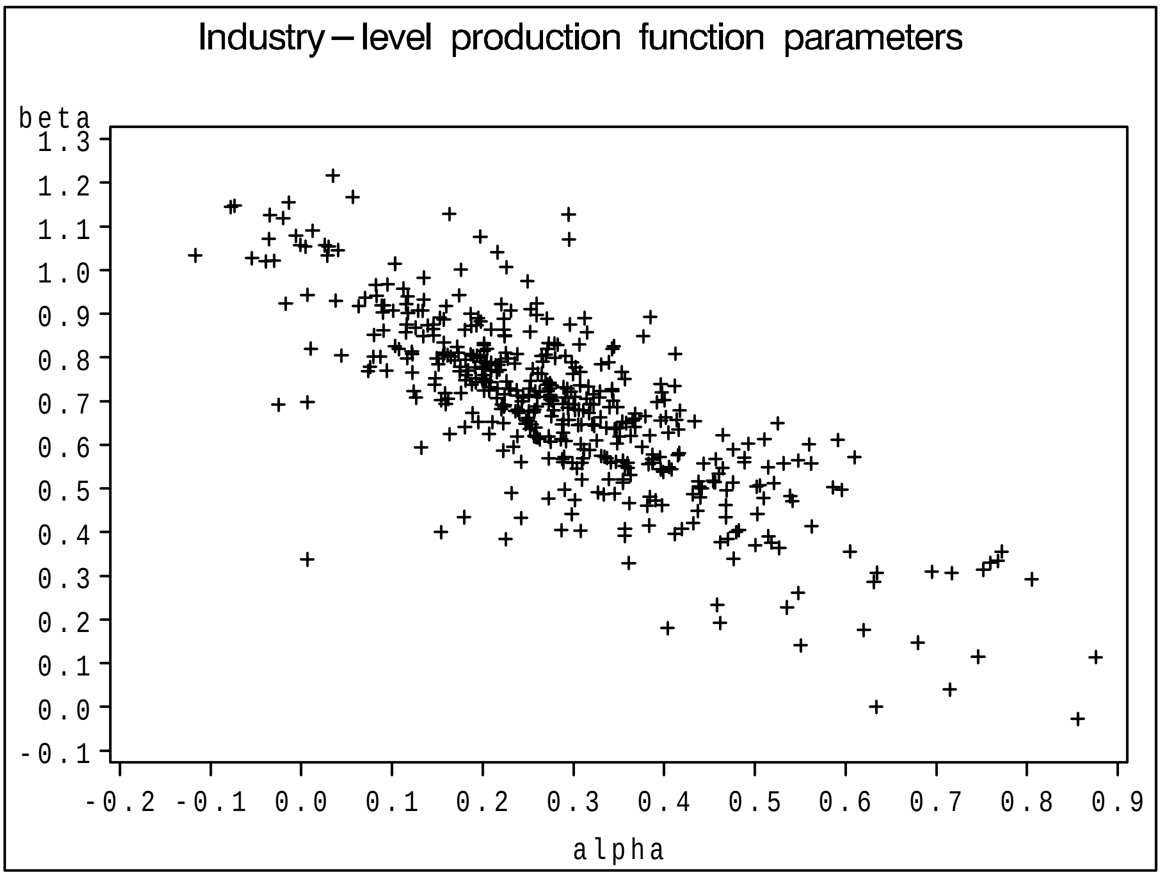

Figure 1: Industry-specific production parameters α : capital share; β : labor share

shipments in our sample. The bad news is that smaller firms are disproportionately under-represented. The reason is that for plants with fewer than 250 employees, the ASM samples plants with a probability that is an increasing function of the plant's size (plants with more than 250 employees are sampled with certainty). Smaller plants in one 5-year panel of the ASM are not eligible for the next 5-year sample. Thus these smaller plants are periodically in and out of the ASM sample. We cannot obtain precise estimates for all three parameters of the $AR(1)$ for plants with few observations. However, we do estimate the plant fixed effects, $\widetilde{\log(z_j)}$, for the entire sample.

Table 1 presents our estimates of the means and standard deviations of the $AR(1)$ parameters.

For each parameter, we present an "unweighted" and a weighted average and standard deviation. The "unweighted" statistics actually weight each plant by the sum of the plant's ASM weights. Thus plants are given more weight if they have a smaller sampling probability (typically smaller plants) and if they appear in the LRD for more years. For the "weighted" averages and standard deviations, we create tvs-share weights. For each plant-year observation, we multiply the plant's tvs-share in manufacturing by the plant's ASM weight in that year. Then we sum this product across all the plant's observations. This is the plant's weight.

The first column of table 1 shows the average across plants of within-plant mean log productivity. These are the averages of the $\widetilde{\log(z_j)}$'s estimated in equation (3). For ease of interpretation, in column 2 we present the mean across plants of the plant's mean productivity in levels. Comparing the first two rows of the table, we can see that on average larger plants are more productive than smaller plants. The last two rows of table 1 present our "unweighted" and weighted averages based on the smaller sample of plants with > 15 observations. Recalling that these plants tend to be larger, the first two columns again lead us to conclude that big plants are more productive than

Table 1: Means and standard deviation for the $AR(1)$ parameters

	$\widetilde{\log(z_J)}$	$\tilde{z}_J$	ρ_j	σ_j	$\frac{\sigma_j}{\widetilde{\log(z_J)}}$
Whole sample - No weights	-0.228	1.02	-	-	-
N= 191,200 plants	(0.58)	(1.00)			
Whole sample - Weighted	-0.02	1.27	-	-	-
	(0.48)	(1.14)			
Small sample - No weights	-0.086	1.124	0.373	0.397	-1.70
N = 19,552 plants	(0.42)	(0.65)	(0.30)	(0.25)	(212.5)
Small sample - Weighted	0.008	1.303	0.406	0.445	1.04
	(0.44)	(0.74)	(0.30)	(0.27)	(122.3)

Small sample: Plants with > 15 observations.
Weights: Shares in total manufacturing value added.

small ones. From the standard deviations in the third and fourth columns, we can see that either there is huge variation in ρ's and σ's across plants, or the small sample variation of our estimators is large. The former conclusion would be consistent with the existing literature which tends to find a lot of plant-level heterogeneity. Comparing the unweighted estimates to the weighted estimates, plant size does not seem to matter much for the variation in our estimates of ρ's and σ.

Perhaps the most important (and robust) finding in table 1 is that average persistence of the plant-specific shocks is surprisingly low. The unweighted average ρ is only 0.37 and the weighted average is only 0.40. Estimates of the persistence of productivity shocks in aggregate data tend to be much larger, as we will see in a moment.

To get an idea of the magnitude of the plant-specific shocks, we take our estimates of each plant's σ and divide it by that plant's log fixed effect, $\widetilde{\log(z_j)}$. The final column of table 1 presents the unweighted and weighted averages of those ratios. The plant-specific shocks are large–the unweighted average percentage is 170% in absolute value.

8 The Role of Aggregation

To show how aggregation can affect estimates of industry and aggregate productivity shocks, we now take a different approach. We estimate an aggregate production function equation analogous to equation (1), but now an observation is a four-digit industry in a given year. That is, we first aggregate our data up to the industry level and then estimate the following equation:

$$\log(y_{it}) = \log(A_{it}) + \alpha \log(K_{it}) + \beta \log(L_{it}) + \sum_{t=1977}^{1999} \iota_t I_t \tag{5}$$

Obviously in equation (5) the production function parameters are restricted to be the same across all industries. The residuals A_{it} are industry shocks. They are closely related to η_{it}, the industry shocks from the plant-level regressions, but the A_{it}'s contain some within-industry reallocation effects. Compared to the plant-level shocks z, the A_{it}'s are more persistent and vary less across observations.

The ι_t's, the coefficients of year dummies in (5), are estimates of aggregate shocks. These are closely related to θ_t, the aggregate shocks from the plant-level regressions, but the aggregate equation, (5), does not control for within-industry reallocation. The ι_t's are biased estimates of θ_t.

We compare industry shocks estimated from aggregate data (the A_{it}'s) to our plant-level shocks (the z's). Using the residuals from the aggregate production function equation, equation (5), we estimate an AR(1). Table 2 shows the results. In the first row (labeled "unweighted"), we present weighted least squares estimates, where the weight for each aggregate observation is the sum of the ASM weights of the plants used to compute that industry-year observation. For the "weighted" estimates in the second row, the weights are the sums of the tvs-shares of the plants. Comparing these estimates to our averages from the plant-level regressions, we find that aggregated industry shocks

Table 2: Persistence and variance of industry-aggregate and plant-level shocks

Shock type	ρ	σ
Industry Aggregate - unweighted	0.81 (0.01)	0.22
Industry Aggregate - weighted	0.92 (0.00)	0.38
Plant-level (z)- unweighted	0.37 (0.30)	0.40 (0.25)
Plant-level (z)- weighted	0.40 (0.30)	0.45 (0.27)

are poor estimates of idiosyncratic plant-level shocks z. They are much more persistent than z; they suggest a smaller magnitude of shocks (σ); and obviously they hide a huge amount of plant-level heterogeneity.

9 Robustness Checks

9.1 Lumpy Investment and Job Flows

In recent years researchers have found that adjustments of employment and capital at the plant level are lumpy.[4] If these lumpy adjustments are the result of (non-convex) adjustment costs, then not taking account of them may bias our estimates of productivity and the persistence and variability of productivity at the plant level. We augment our reduced-form production function model with measures of investment and job flow spikes.

Lumpy investment measures. Following Power (1998), Sakellaris (2004), and others, we measure investment spikes using equipment investment, equipment capital stocks and various definitions of an investment spike. Define the gross investment rate, GINVRATE, as NMREAL/KEQ, new investment in equipment divided by the equipment capital stock. We create measures of both relative and absolute investment spikes. A

[4]See Ricardo Caballero and Haltiwanger (1997) on employment, and Doms and Dunne (1998) on capital.

relative investment spike occurs in a given year if the gross investment rate is greater than γ times the plant's median gross investment rate, where γ is some (arbitrary) threshold. For sensitivity analysis, we set γ equal to 1.75, 2.5, and 3.25. We also define an alternative measure of relative investment spike as a year in which the plant's gross investment rate is greater than two standard deviations above its median investment rate. We define an absolute investment spike as a year in which the plant's gross investment rate is greater than 20%. Finally, we define a multi-year investment spike as 2 or more consecutive years in which the plant's gross investment rate is greater than ω times the plant's median gross investment rate. Again following Power (1998), we choose (arbitrary) values of 0.80, 0.85, and 0.90 for ω. We will define the benchmark definition of a investment spike as the occurance of an absolute investment *or* a multi-year investment spike with $\omega = 0.90$.

Lumpy job flow measures. Again following Sakellaris (2004) we define a burst of job creation as a year in which the the growth rate of plant hours is greater than 10% and the previous year's growth rate of plant hours is less than 10% in absolute value. We define a burst of job destruction as a year in which the plant's growth rate of plant hours is less than -10% and the previous year's growth rate of plant hours is less than 10% in absolute value.

An investment or job flow spike may be related to productivity not just in the period (or periods) in which it occurs, but also in prior or subsequent periods.[5] To account for this possibility, for each spike measure we also include dummy variables for 1 and 2 years before and after the spikes. For each alternative measure of investment spikes, we estimate equation (4) modified to include a full set of investment spike and job flow spike dummies. The first stage estimation results changed very little with the addition of spike dummy variables. For example, for the benchmark specification the shipments-share-weighted average estimates of α and β are 0.29 and 0.68, respectively.

[5] Sakellaris (2004) provides evidence that this is the case.

Table 3: Means and standard deviations for the $AR(1)$ parameters, with spikes in 1st stage

	$\widetilde{\log(z_J)}$	$\tilde{z}_J$	ρ_j	σ_j	$\frac{\sigma_j}{\widetilde{\log(z_J)}}$
Large sample - No weights	-0.226	1.022	-	-	-
N = 191,200 plants	(0.57)	(0.99)			
Large sample - Weighted	-0.029	1.265	-	-	-
	(0.47)	(1.14)			
Small sample - No weights	-0.083	1.125	0.365	0.401	-12.90
N = 19,552 plants	(0.41)	(0.64)	(0.32)	(0.25)	(1527.3)
Small sample - Weighted	0.007	1.286	0.391	0.445	-12.02
	(0.43)	(0.69)	(0.31)	(0.26)	(799.1)

Small sample: Plants with > 15 observations.
Weights: Shares in total manufacturing value added.

As before, we take the residuals from the production function regressions and run plant-level AR(1)'s to estimate the plant-level fixed effects and the persistence and standard deviation of the plant-specific shocks. Table 3 presents means and standard deviations of our estimates of the AR(1) parameters, where we used our benchmark definition of an investment spike in the first stage of estimation.[6] Compared to table 1 our estimates of the fixed effects change somewhat, but the average persistence and standard deviation of the plant-specific shocks change very little. We conclude that the low level of persistence and high level of variance in the plant-specific shocks is not caused by not taking account of spikes in investment, job creation or job destruction.

9.2 Missing observations.

The rotating 5-year panel structure of the ASM causes many plants to have missing observations (see the data appendix for details on the ASM panel structure).[7] To reduce

[6]Estimates of ρ and σ using the other measures of investment spikes do not differ substantially from these. The other estimates are available from the authors upon request.

[7]It is possible that some smaller plants may have missing observations because of genuine shutdown and re-entry, but the majority of missing observations seem to be caused by the panel structure of the ASM.

Table 4: Means and standard deviations for the $AR(1)$ parameters, no missing years

	$\widetilde{\log(z_J)}$	$\widetilde{z}_J$	ρ_j	σ_j	$\frac{\sigma_j}{\widetilde{\log(z_J)}}$
Large sample - No weights	-0.242	1.016	-	-	-
N = 144,178 plants	(0.61)	(1.06)			
Large sample - Weighted	-0.006	1.299	-	-	-
	(0.48)	(1.27)			
Small sample - No weights	-0.066	1.141	0.383	0.387	-4.622
N = 9,636 plants	(0.42)	(0.63)	(0.32)	(0.24)	(224.3)
Small sample - Weighted	0.028	1.312	0.402	0.437	-9.73
	(0.43)	(0.71)	(0.31)	(0.25)	(376.3)

Small sample: Plants with > 15 observations.
Weights: Shares in total manufacturing value added.

the possibility that our finding of low persistence of the idiosyncratic shocks is driven by missing observations, we take the residuals from our benchmark specification with investment and job flow spikes and we throw out plants with any observations missing between the first year the plant is observed and the last year the plant is observed. Table 4 presents our estimates for the means and standard deviations of the AR(1) parameters for this sample. The estimates of the persistence of idiosyncratic productivity shocks differ very little from the estimates presented in table 3. Thus we conclude that our finding of low persistence is not driven by missing observations.

9.3 The Importance of Fixed Effects

Plant fixed effects are an important reason for our estimates of low average persistence of plant-specific productivity shocks. To show this, we take the residuals used to estimate the results in table 4, and we again estimate plant-specific AR(1)'s, this time without plant fixed effects. Table 5 shows the results. The average estimated standard deviation of the shocks is about the same, but the average estimated persistence increases from about 0.40 in the presence of plant fixed effects to about 0.59 when there are no plant fixed effects.

Table 5: Persistence and variance of plant-level shocks without fixed effects

Shock type	ρ	σ
Without fixed effects - unweighted	0.590	0.413
N = 9,636 plants	(0.33)	(0.25)
Without fixed effects - weighted	0.590	0.461
	(0.29)	(0.26)

10 Conclusions

In line with previous research we find a tremendous amount of heterogeneity among manufacturing plants, both in terms of productivity and in terms of the persistence and variability of productivity at the plant level. We find that on average the persistence of plant-specific productivity shocks is surprisingly low. These estimates of low persistence are robust to the inclusion of measures of lumpiness in investment and job creation and destruction, different measures of the capital stock, and the exclusion or inclusion of plants with missing observations. One caveat to our finding of low plant-specific persistence is that plant-level fixed effects seem to be important. Under the assumption of no plant fixed effects, the yearly autocorrelation of plant-specific productivity shocks increases from about 0.37 to about 0.59 on an annual basis. Note, however, that when we use the same data aggregated to the 4-digit industry level, we estimate autocorrelations from 0.81 to 0.92. Thus estimates of the persistence of productivity shocks from aggregate data are upward-biased estimates of the persistence of plant-specific productivity shocks.

11 Next Steps and Ideas for Future Research

- Hurwicz bias: our estimates of the autocorrelation of productivity may be downward-biased because of the small sample properties of our OLS estimator. Blundell and Bond (1998, Journal of Econometrics) may provide a remedy for this.

- We need to deal with endogeneity of inputs, especially of labor input. In some industries selection bias from exit may also be a problem. We can do this by applying the Olley and Pakes (1996) methodology, at least for some industries. Another possibility for dealing with the endogeneity of inputs is applying Blundell and Bond (1998).

- Variance decomposition: what percentage of the total variation of value added is explained by the variation of each type of shock (aggregate, industry, plant-specific)?

- The dynamics of relative weights: how does the relative importance of the different types of shocks change over time?

- Build a GE model of between- and within-industry reallocation and use the data set to test the model.

A Data Appendix

We use the Census Bureau's Longitudinal Research Database (LRD) and the Bureau's Longitudinal Business Database (LBD). The LBD currently runs from 1976 to 1999 and is described in Jarmin and Miranda (2002). The LRD consists of the Census of Manufactures (CM), which is conducted in years ending in 2 and 5, and the Annual Survey of Manufactures (ASM), which is conducted in non-census years. For deflators and depreciation rates we use the dataset available on John Haltiwanger's anonymous ftp site: ftp://haltiwan.econ.umd.edu.

Production-worker equivalent hours. We construct our hours measure as follows. PLANT-HOURS, WORKER-WAGES, and SALARIES-AND-WAGES are observed. PLANT-HOURS is total production work hours at a plant. WORKER-WAGES is the plant's total production worker wages. PRODUCTION-WAGE = WORKER-WAGES/PLANT-HOURS thus gives the average production worker wage rate at the plant. SALARIES-AND-WAGES is total salaries and wages at the plant–this is the sum of production workers wages and non-production worker salaries. PW-HOURS = SALARIES-AND-WAGES/PRODUCTION-WAGE gives the total production-worker-equivalent hours for the plant. We take the log of PW-HOURS as our measure of labor input. None of the plants in our samples have missing values for any of the variables we use to construct our labor input measure. However, PLANT-HOURS are zero for between 400 and 1600 plants in each year. WORKER-WAGES are zero for between 400 and 1700 plants in each year. SALARIES-AND-WAGES are less likely to be zero; the number of plants with zero SALARIES-AND-WAGES ranges from well below 100 in the lowest year to roughly 1000 in the highest year.

Capital stock. We create our capital stock measure using the perpetual inventory method. For most years, the LRD keeps track of capital and investment measures separately for buildings (or structures) and machinery (equipment). For buildings we compute the capital stock as $KST_t = (1-STDPR)KST_{t-1}+NBREAL_t$, where KST_t is the building capital stock in year t; STDPR is the 2-digit-SIC depreciation rate for buildings from Haltiwanger's dataset; and NB_t is real investment in new buildings in year t. To compute real investment, we use the current-dollar investment in buildings, NB, and the 2-digit-SIC investment deflator from Haltiwanger's dataset, PIINVS: NBREAL = NB/PIINVS. Similarly, for the machinery capital stock we have $KEQ_t = (1 - EQDPR)KEQ_{t-1} + NMREAL_t$, where EQDPR is the 2-digit depreciation rate for machinery (equipment); NMREAL is real investment in machinery; and NMREAL=NM/PIINVE, where PIINVE is Haltiwanger'S 2-digit investment deflator for machinery (equipment). Following Haltiwanger, we compute the real initial capital capital stocks for machinery and buildings as $KEQ_{initial} = BAE * (NKCEQ/GKHEQ)/PIINVE96$ and $KES_{initial} = MAE * (NKCST/GKHST)/PIINVS96$, where: MAE and BAE are the nominal book values at the end of the year for machinery and buildings, respectively; NKCEQ and

NKCST are the 2-digit industry level nominal net capital stocks for equipment and structures, respectively; GKHEQ and GKHST are the 2-digit industry level real gross capital stocks for equipment and structures, respectively, in 1996 dollars; PIINVE96 and PIINVS96 are Haltiwanger's 2-digit investment deflators for equipment and structures in 1996. There are some problems in calculating capital stocks in the later years of the LRD. BAE and MAE are reported in the Annual Survey every year up to and including 1985 and in 1992. In the 1997 Census, the Annual Survey questionnaire only asked about total assets at the end of the year, i.e., the sum of building assets and machinery assets. In 1986, 1988-1991, 1993-1996, and 1998-9, no questions about assets were asked in the Annual Surveys. However, questions about capital expenditures were asked in all years. For plants that enter between 1986 and 1996 and continue through a Census year, we can use a backwards perpetual inventory method to compute capital stocks. So, for example, for a plant that appears in the LRD in 1989, 1990, 1991, 1992, and 1993, first we initialize the capital stocks for equipment and structures in 1992: $KEQ_{92} = BAE_{92} * (NKCEQ_{92}/GKHEQ_{92})/PIINVE96$ and $KES_{92} = MAE_{92} * (NKCST_{92}/GKHST_{92})/PIINVS96$. Then we use the capital expenditure data to compute capital stocks for the other years, first going backwards: $KEQ_{t-1} = (KEQ_t - NMREAL_t)/(1 - EQDPR_t)$ and $KST_{t-1} = (KST_t - NBREAL_t)/(1 - STDPR_t)$; and then forwards: $KEQ_t = (1 - EQDPR_t)KEQ_{t-1} + NMREAL_t$ and $KST_t = (1 - STDPR_t)KST_{t-1} + NBREAL_t$. We use the same procedure for plants that enter the LRD in 1986, except that we initialize the capital stock using the 1987 asset variables. For plants that enter the LRD after 1992, we use an analogous procedure, but we can only compute the total initial real capital stock using the total assets variable. We compute the 1997 real capital stock for these plants as $K_{97} = TAE_{97} * (NKCEQ_{97}/GKHEQ_{97})/PIINVE96$. To calculate the rest of the capital stock series, we use the initial real capital stock, real total capital expenditures, and the equipment depreciation rates, again going forwards: $K_t = (1 - EQDPR)K_{t-1} + TCEREAL_t$, where $TCEREAL = TCE/PIINVE$; and backwards: $K_{t-1} = (K_t - TCEREAL_t)/(1 - EQDPR_t)$.

There are three sets of plants for which we cannot compute capital stocks from plant-specific data: (i) plants that enter the LRD in 1988-1991 and exit the LRD before 1992 (roughly 12,000 plants); (ii) plants that enter the LRD in 1993-1996 and exit the LRD before 1997 (about 42,000 plants); (iii) plants that enter the LRD in 1998 or 1999 (roughly 20,000 plants). We drop these three sets of plants from our sample. We would also be unable to compute plant-specific capital stocks for plants that both enter and exit the LRD in 1986, but no plants in our sample fall into this category.

In each year of our sample, there are some plants for which we compute an initial capital stock of zero or even less than zero. Because of the log specification of our production function we cannot use these observations. In years for which we have initial asset values (i.e., in census years and in ASM years before 1986), our capital stock measure is zero because the asset values report in the ASM were zero. The number of initial capital stocks measured at zero varies by year. In 1976 and 1977 there are between 500 and 700 in each year. In 1978-9, 1981, and 1983-88 there are fewer than 100 such observations in each year. In 1980, 1982, and 1997 there are between 100 and 200 each year. In 1990-92, and 1995-6 there are between 200 and 300 each year. In 1993 and 1994 the number of initial capital stocks measured at less than or equal to zero jumps to about 900 and about 1500, respectively. Recall that for all plants that enter the LRD between 1993 and 1996, the initial capital stock is computed starting from total assets in 1997 and using a backwards perpetual inventory method. For these plants we are using the equipment (machinery) depreciation rates, which are higher than the building (structures) depreciation rates. Thus it seems unlikely that the negative initial stock measures are caused by our depreciation rates. All of these observations (although not necessarily all observations for these plants) are thrown out of our sample.

The ASM and CM also collect data on *used* machinery expenditures, used building expenditures, machinery and building retirements, and rental payments for buildings and machinery. Olley and Pakes (1996) note that used capital is more important for small plants. Used capital expenditures are also not available in the LRD after 1996. Retirements for capital are not available in the LRD for 1991 or any year after 1992. Rental payments for capital are also unavailable in the LRD for 1991 or any year after 1992 except 1997. To be consistent across our sample years, we follow most studies that use the LRD and do not use these variables to compute our capital stock measure.

Value-added. Our measure of value-added is the log of the real value of shipments minus the real cost of materials plus the real change in the inventories. The current dollar total value of shipments, TVS, is observed in the LRD. The shipments deflator, PISHIP, and the materials deflator, PIMAT in Haltiwanger's dataset, are at the 4-digit SIC level and are the same as the materials and shipments deflators in the NBER/CES productivity database. For the real value of shipments, we compute $TVS - REAL = TVS/PISHIP$. For the real cost of materials, MATQ, we calculate $MATQ = CM/PIMAT$. In the LRD we also observe the plant's total inventories at the beginning of the year, TIB, and at the end of the year, TIE. We take the difference of these two and deflate by the shipments deflator to arrive at the real total change in inventories: $TOT_CHG_INV_REAL = (TIE - TIB)/PISHIP$. We then calculate real value added as $TVS_REAL - MATQ + TOT_CHG_INV_REAL$.

Plant age. We measure plant age as the year minus the year that the plant first appeared in the dataset. Obviously this only puts a lower bound on age for plants that appear in the first year of our sample, 1976.

Industry coding. We assign plants to an industry based on the longitudinal microdata. In the first stage, we are estimating production functions at the 4-digit SIC industry level, and in the second stage we are estimating the dynamics of plant-level productivity. Thus we assign industry codes so that plants do not change 4-digit SIC industries. The Census Bureau makes its greatest effort at accurate plant level industry coding in the census years. There was a major revision of the SIC coding system in 1987. However, in the CM years of the LRD files, the plants have been recoded using the 1987 SIC system *based on the 7-digit products that the plants were producing.* Thus we can use the 1987 SIC coding system and the CM years to create relatively consistent 4-digit SIC industries across time using the microdata. We start with a plant's 4-digit industry code in 1987. If the plant existed in 1987, we assign the plant's industry code in that year to all years for that plant. If the plant did not exist in 1987, we start with 1992, then 1997, then 1982, then 1977. Some plants are not observed in census years and are only observed before 1987. For these plants if the plant's existing SIC code is the same as a 1987 SIC code, then we assign the existing code as the 1987 SIC code. We assign 1987 sic codes to roughly 23,000 observations this way. This leaves about 3700 observations to which we are not able to assign a 1987 SIC code. These plants are dropped from our sample. After assigning 1987 SIC codes, we drop plant-year observations for which the industry code is not in manufacturing (SIC=2000-3999). This reduces the sample by about 7,000 observations.

Our industry coding scheme gives us 460 industries. Thus in the first stage of estimation we estimate 460 different production functions. Seven of these industries have so few observations in most or all years that the matrix of regressors is not full rank and/or they do not pass the Census Bureau's disclosure rules for certain years.[8] We drop these industries and observations from our sample. This reduces the sample size by 3,163 observations, and leaves us with 453 industries. One of the remaining industries (SIC

[8]The industries are: 2067 - chewing gum; 2097 manufactured ice; 2386 leather and sheep-lined clothing; 2429 special product sawmills, not elsewhere classified; 2771 greeting cards; 2895 carbon black; and 3962 artificial flowers. The latter industry is a 1977 SIC industry and does not exist in the 1987 SIC system.

code 2721, periodicals: publishing and printing) has hundreds of observations in each year from 1976 to 1998 but no observations in our sample in 1999. For this industry we run the regression in equation (4) without the 1999 year dummy.

Further description of samples. About a third of plants in the LRD have 5 or fewer employees. For these plants, the data is imputed from Administrative Records (AR). We follow most researchers who have used the LRD and drop these AR plants from our sample. Part of our goal in this research is to compare the persistence and variability of plant-level idiosyncratic productivity shocks to aggregate productivity shocks. As a result, we would like our aggregate measures to be comparable across time. In the LRD, ASM years have about 55,000 to 75,000 plants and CM years in our sample have about 350,000 to 400,000 plants. However, plants that are not sampled with certainty in the ASM years have weights based on their sampling probability, and these plants also have these ASM weights in the CM years. Thus to make our aggregate measures comparable across years we use only plants that have positive, non-missing ASM-weights.

To create the best possible longitudinal links, we link our LRD plants to the LBD. The LBD covers over 23 million unique establishments, including all industries covered by the Economic Census (for a more detailed description of the LBD, see Jarmin and Miranda, 2002). In most years, including only observations that are in both the LRD and the LBD reduces the sample by 3,000 to 4,000 plants per year. There are about 1.4 million plant-year observations in this combined LRD-LBD sample of non-AR plants with positive, non-missing ASM weights. Throwing out plants for which we cannot compute capital stocks (see the capital stock section above) reduces the sample further to about 1.3 million observations. We call this our "Whole" Sample, which we use to estimate the production functions and one set of plant fixed effects (see tables 1 and 3). To estimate the persistence and standard deviations of the plant-specific productivity shocks, we limit the sample to plants with more than 15 observations. This reduces our "Small Sample" size to 403,980 observations.

Many plants, especially plants in our "Whole Sample" have years missing between the first and last year the plant is observed. About 85% of the plants with years missing have years missing because of the structure of the ASM panel. There are two scenarios: (1) these plants appear in an ASM panel two years after appearing in a Census of Manufacturers; or (2) they appear in one 5-year ASM panel, disappear in the next one, two, or three panels, and reappear in another panel. We keep these plants in our sample when we estimate the production function parameters in the first stage. For run the second stage of estimation on sample both with and without the plants with missing observations.

First we run the regressions on the sample that includes plants with missing observations (see results in tables 1 and 3). However, we want to reduce the possibility that our result of low persistence in plant level idiosyncratic productivity shocks is driven by missing years at the plant level. Thus we run another set of second stage regressions, eliminating from our samples all plants that have at least one year missing between the first observation of the plant and the last observation of the plant. This reduces our large sample substantially, from 1.3 million plant-year observations to a little over 800,000 plant-year observations.

The additional requirement that the plant be observed for more than 15 years reduces the "Small Sample" to 209,518 plant-year observations for 9,636 plants. Throwing out all the plants with missing years between the plant's first and last observation changes our estimate of ρ, but not by much: compare tables 3 and 4. [In future work, we may test the sensitivity of our results to different rules for excluding or including plants with missing years.]

References

Bartelsman, Eric and Mark Doms, "Understanding Productivity: Lessons from

Longitudinal Microdata," *Journal of Economic Literature*, 2000.

Blundell, Richard and Stephen Bond, "Initial Conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, 1998.

Caballero, Eduardo Engel Ricardo and John Haltiwanger, "Aggregate Employment Dynamics: Building from Microeconomic Evidence," *American Economic Review*, 1997, *87* (1), 115–137.

Doms, Mark and Timothy Dunne, "Capital Adjustment Patterns in Manufacturing Plants," *Review of Economic Dynamics*, 1998.

Dwyer, Douglad, "Are Fixed Effects Fixed? Persistence in Plant Level Productivity," 1996. CES Working Paper.

Jarmin, Ron and Javier Miranda, "The Longitudinal Business Database," 2002. CES Working Paper.

Jovanovic, Boyan, "Selection and the Evolution of Industry," *Econometrica*, May 1982, *50.*

Levinsohn, James and Amil Petrin, "When Industries Become More Productive, Do Firms? Investigating Productivity Dynamics," 1999. Working Paper.

Olly, Stephen and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, November 1996.

Power, Laura, "The Missing Link: Technology, Investment, and Productivity," *The Review of Economic and Statistics*, May 1998.

Sakellaris, Plutarchos, "Patterns of Plant Adjustment," *Journal of Monetary Economics*, 2004.

Storesletten, Kjetil, Chris Telmer, and Amir Yaron, "Asset Pricing with Idiosyncratic Risk and Overlapping Generations," 1999. mimeo, Carnegie Mellon University.